

---

## Chapter 5

# Retrospective Analysis of Passage Mortality of Spring Chinook of the Columbia River

by Rick Deriso, David Marmorek and Ian Parnell

### Abstract

There have been decades of debates about passage mortality in the Columbia River. We empirically estimated instantaneous in-river passage mortality ( $m$ ) and its associated probability distribution, using spawning and recruitment data for seven Snake River and six lower Columbia River spring chinook populations. Our empirical estimates of  $\mu$  showed low bias, and were generally close to (and between) those produced by the two more mechanistic models CRiSP and FLUSH, though significantly higher than one version of CRiSP. A total of 37 models were applied to the data, incorporating different assumptions about spawner-measurement error, transport survival, intrinsic productivity, methods of estimating  $m$  and 'year effects' which accommodate common factors affecting the survival of all stocks. The 12 'top' models estimates of mean  $m$  ranged from 0.55-1.90 (grand mean of 1.09); two of these models simply made  $m$  proportional to Water Transit Time (WTT). These estimates of  $\mu$  imply that in-river passage from Lower Granite to John Day dam reduced recruitment, of 1970-1990 broods by an average of 42-85% (grand mean of 66%). The time trend in  $\mu$  was cyclical and generally moderate to high for 1970's brood years, low for 1980-1983, slightly cyclical around the long-term average for 1984-1989, and high in 1990. The year effect shifts from generally positive effects on 1952-1968 brood years to generally negative effects on 1970-1990 broods. Year effects were not correlated with WTT,  $m$ , the North Pacific Index (NPI) or an Upwelling Index, but were inversely correlated with Bristol Bay (Alaska) sockeye survival anomalies. We discuss the improvements required to apply these models to prospective analyses.

### 5.1 Introduction

Stock and recruitment data can provide a powerful basis for inferential hypothesis testing and estimation regarding mortality experienced by salmon during their down-river migration. In order to develop inferences we need a simple model of the salmon populations of the Columbia River system that can be used to simultaneously estimate factors affecting the survival of each salmon population.

Stock and recruitment models for the chinook stocks of the Columbia River contain four basic components: natural mortality, fishing mortality, in-river passage mortality, and spawning. In the models described here we focus on a series of simple models that assume that survival of chinook from rivers of the Columbia share many similar aspects. In particular, ocean mortality rates are assumed to share many common attributes across stocks from different streams. By making such an assumption, the primary systematic difference in survival of chinook from up-river, mid-river, and down-river streams can be attributed to differences in their in-river passage mortality, after accounting for basic population differences in their intrinsic productivity (that is, the Ricker  $a$  and  $b$  parameters) and individual stream random effects. The model contains a 'year-effect' component to mortalities which permits the model flexibility to accommodate a common factor affecting survival of all the stocks (as for example, a

common ocean mortality factor). We exploit the fact that many dams along the Columbia and Snake River were initiated at various times encompassed by the time series of available stock and recruitment data and that various populations are exposed to different numbers of dams, much in the spirit of the staircase experimental design method of Walters et al (1989).

The goal is to estimate in-river passage mortality and its associated probability distribution by fitting stock and recruitment models to available data on recruitment and escapement (plus any ancillary information available). In particular we estimate the passage mortality for spring Chinook salmon that they experience during their seaward journey from sub-basins of the Snake River through a total of five main-stem dams down to the John Day dam located in the mid-Columbia River; the salmon pass through three additional main-stem dams from the John Day dam to the ocean. We utilize the population time series available for seven of the Snake River sub-basin populations and six population time series available in the mid- and lower-river sub-basins (described more fully in the next Section). Recruitment is estimated for the mouth of the Columbia River based on the methods described in Beamesderfer et al (1996), and summarized in Section 3.3 of Chapter 3. We compare a series of alternative models including ones that use estimates made by two passage models (CRiSP - Anderson et al. 1996<sup>1</sup>, and FLUSH - Wilson et al. 1994) and characterize differences and similarities between the estimates. A prime motivation for this analysis was to assess to what degree competing models' more mechanistic approaches to estimating passage mortality agree with the more hybrid/empirical models we used.

We acknowledge with much thanks the contributions of stock-recruitment data from Howard Schaller and Ray Beamesderfer (ODFW), Charlie Petrosky (IDFG), and Olaf Langness (WDFW); FLUSH model projections from Earl Weber (CRITFC) and Paul Wilson (CBFWA); CRiSP model projections from Jim Anderson and Josh Hayes (Columbia Basin Research) and David Askren (BPA); and spawner-measurement error estimates from Charlie Petrosky (Appendix 5-1). Many PATH participants contributed towards the formulation of this approach, through a series of workshops and technical meetings.

## **5.2 Spring and Summer Chinook Salmon Populations Examined**

Thirteen populations of Chinook salmon were analyzed in this study. They represent three down-river subbasins — those of the Wind River, Klickitat River, and Warm Springs River; three populations in the John Day subbasin system — the John Day Main-stem, John Day Middle Fork, and John Day North Fork; and seven up-river subbasins all branching from the Snake River — those of the Imnaha, Minam, Bear Valley, Marsh Creek, Sulphur Creek, Poverty Flat, and Johnson Creek Rivers. Those thirteen populations represent the total number of populations on the lower to middle Columbia River system and Snake River system for which time series of spawner and recruitment information were available. Additional population time series are available for the Upper Columbia River, which we plan to analyze in a future report. Table 5-1 summarizes the number of main-river dams located below each river subbasin along with the number of years of spawner and recruitment information available.

Each of these subbasins are described in Chapter 3 of this report (Schaller et al. 1996), and in more detail in Beamesderfer et al. (1996).

---

<sup>1</sup> A description of the theory, calibration, and validation of the CRiSP model can be found in Anderson et al (1996). Documentation of parameters used for the CRiSP survival values are identified in (Askren 1996). Several sensitivity analyses for CRiSP survival estimates are also provided in Anderson et al (1996, p.192 - 201) and Askren (1996).

**Table 5-1:** Summary information on the thirteen chinook populations analyzed in this study.

Sub-basin	Brood Years of Paired Spawner-recruit Data	Number of Main-stem Dams Below Sub-basin
1. Wind	1973 - 1990	1
2. Klickitat	1966 - 1990	1
3. Warm Springs	1969 - 1990	2
4. John Day Mainstem	1959 - 1990	3
5. John Day Mid Fork	1959 - 1990	3
6. John Day North Fork	1959 - 1990	3
7. Imnaha	1952 - 1990	8
8. Minam	1954 - 1990	8
9. Bear Valley	1957 - 1990	8
10. Marsh Creek	1957 - 1990	8
11. Sulphur Creek	1957 - 1990	8
12. Poverty Flat	1957 - 1990	8
13. Johnson Creek	1957 - 1990	8

### 5.3 Population Models

The general model structure is based on a Ricker type spawner-recruitment model, similar in structure to the Ricker models used in Chapters 3, 4, and 9, but explicitly accounting for measurement error:

$$y_{t,i} = x_{t,i} + a_{t,i} - b_i e^{x_{t,i}} - m_{t,i} + \mathbf{e}_{t,i} \quad [1]$$

$$\ln(S_{t,i}) = x_{t,i} + \mathbf{e}'_{t,i} \quad [1a]$$

Where:

- $y_{t,i}$  =  $\ln R_{t,i}$
- $R_{t,i}$  = Columbia River “observed” returns (recruitment) originating from spawning in year  $t$  and river sub-basin  $i$  (# of fish)
- $S_{t,i}$  = “observed” spawning in year  $t$  and river sub-basin  $i$  (# of fish)
- $x_{t,i}$  = “true” logarithm of spawning
- $a_{t,i}$  = Ricker  $a$  parameter, which depends on year and river (unitless)
- $b_i$  = Ricker  $b$  parameter, which depends on river (1/fish)
- $m_{t,i}$  = in-river passage mortality which depends on year and river (unitless)
- $\mathbf{e}_{t,i}$  = normally distributed mixed process error and recruitment measurement error term  $N(0, \sigma^2_{\mathbf{e}})$  [ $\ln(\#$  of fish)]
- $\mathbf{e}'_{t,i}$  = normally distributed spawner enumeration measurement error term  $N(0, \sigma^2_{\mathbf{e}'})$  [ $\ln(\#$  of fish)]
- $t$  = year
- $i$  = sub-basin

Notice that model (1) combines all non-dam, density-independent mortality sources into a single parameter  $a_{t,i}$  except for the additive process error  $\mathbf{e}_{t,i}$ . Thus, the  $a$  parameter is composed of a sum of egg-to-fry, fry-to-smolt, and smolt-to-adult density-independent mortality. The density-dependent parameter  $b$  is chosen to be area-specific to reflect the different carrying capacities of different areas. Typically estimates of the  $a$  and  $b$  coefficients are confounded so that we did not attempt to model temporal changes

in carrying capacity explicitly. Note that setting  $e\phi_{t,i}$  and  $m_{t,i}$  to zero, and substituting  $\ln(S_{t,i})$  for  $x_{t,i}$  in Equation [1] generates the logarithmic form of the Ricker model.

Chinook salmon may return to spawn at age 3, 4 or 5. The model does not contain separate mortality terms for each age class of a given brood. Instead the assumption was made of a constant post-age-4 ocean survival. Inter-annual variation in ocean survival is confined to the fish's first two years of life in the ocean and estuarine environment. In their final years at sea, ocean mortality for Chinook is probably lower —  $M=0.2$  per year was estimated in Ricker (1976) — and therefore likely does not induce large systematic variations in total ocean mortality of Chinook, as compared to total mortality rates on the order of 4.07 (Bradford 1995), which covers total mortality from the smolt life-stage to adult return for similar stream-type chinook. Other assumptions regarding ocean survival are discussed further below. The data assumption, namely that stock and recruitment estimates are available by run, utilizes the estimates obtained in Petrosky et al (1995) and recent unpublished estimates obtained from PATH participants.

Structure for the  $a$  parameters was added in three alternative ways:

$$a_{t,i} = a + \mathbf{d}_t \quad [2]$$

$$a_{t,i} = a_i + \mathbf{d}_t \quad [3]$$

$$a_{t,i} = a_i + \mathbf{d}_{t,region} \quad [4]$$

where in each alternative the parameter  $a$  is the Ricker  $a$  parameter;  $i$  is the subbasin (e.g. Imnaha); 'region' refers to either the 7 up-river subbasins or the 6 downriver subbasins. In Equation [2], the model where the Ricker  $a$  value was assumed to be the same for different stocks, we still retained a separate  $a$  coefficient for the Wind River stock because it is an introduced stock (see Chapter 3 of this report). In Equation [3] the model allows the Ricker  $a$  parameter to vary for each stock. The "year-effect" parameter  $\mathbf{d}_t$  was included to allow for year effects common to all stocks; this would include major ocean mortality changes that affect the survival of chinook salmon during the first two years of ocean life, as well as regional changes in terrestrial climate that affect all stocks. Above average climate conditions have  $\delta_t > 0$ , while below average conditions have  $\delta_t < 0$ .

We wish to stress the simplicity of the model given by Equations [1], [1a] and [3] (the main model we use). Eliminating measurement error ( $\epsilon'_{t,i}$  in Equation [1a]) combining Equations [1], [1a], and [3], and rearranging terms generates the following equation:

$$\ln(R_{t,i} / S_{t,i}) = [a_i + \mathbf{d}_t - m_{t,i}] - b_i S_{t,i} + \mathbf{e}_{t,i} \quad [4a]$$

which is of identical structure to the multiple regression models described in Chapter 4, Section 4.3. The four major differences between our models and the ones in Chapter 4 are: 1) the inclusion of spawning enumeration measurement error as an estimated parameter; 2) the specific independent variables (i.e. on the right side of Equation [4a] used to explain variation in  $\ln(R_{t,i} / S_{t,i})$ ); 3) the use of maximum likelihood estimation procedures to estimate variances; and 4) the grouping of stocks and years for the purposes of parameter estimation. Exponentiating both sides of Equation [4a] generates the familiar Ricker equation plus two additional terms for the year and passage mortality effects:

$$R = S e^{a_i - b_i S_{t,i} + \mathbf{e}_{t,i} + \mathbf{d}_t - m_{t,i}} \quad [4b]$$

Results for  $\delta_t$  and  $m_{t,i}$  are generally presented as their actual values. Exponentiating  $\delta_t$  and  $m_{t,i}$  expresses these parameters as fractional factors which in any given year can increase or decrease the recruitment from that expected by the Ricker parameters alone, that is:

$$R = S \left[ e^{a_i - b_i S_{t,i} + \varepsilon_{t,i}} \right] \left[ e^{d_t} \right] \left[ e^{-m_{t,i}} \right] \quad [4c]$$

Results are presented for both the parameters themselves (i.e.  $\delta_t$ ,  $m_{t,i}$ ) and their exponentiated form, since the latter can be more easily understood. For example, the percent mortality (PRR = percent reduction in recruitment) due to in-river passage is computed by:

$$PRR = 100 * \left[ 1 - e^{-m_{t,i}} \right] \quad [4d]$$

Similarly, the percent change in recruitment due to year effects (PYE) can be expressed as:

$$PYE = 100 * \left[ e^{d_t} - 1 \right] \quad [4e]$$

There is considerable interest in spatial and temporal variations in ocean survival (see Chapters 4 and 12). Using either Equation [2] or Equation [3] as part of the overall model in Equation [1] entails the following assumptions with respect to ocean survival: 1) year-to-year variations in climate which effect the ocean survival of all stocks will be picked up by the year effect  $\delta_t$ ; 2) random variations among stocks in ocean survival (due, for example, to different ocean distributions or timing of ocean entry) will be assigned to the term  $\varepsilon_{t,i}$  in Equation [1], which picks up any unexplained natural variation in recruitment, as well as recruitment measurement error; and 3) variations among stocks in ocean survival are not systematic differences between up-river and mid-lower stocks, as regional stock groups.

To examine the sensitivity of the model to the third assumption, we added equation [4] which is a generalization of Equation [3] to allow for year-effects that differ between the up-river and mid-to-lower river stocks. When equation [4] is implemented, an additional sum-of-squares term is needed in the estimation procedure to characterize the correlation between the year-effects in the two stock groups. If the within-region differences among stocks in year effects are as large as the among-region differences (i.e. no systematic, significant differences between up-river and mid-to-lower river regions), then these 'random effects' will be assigned to the residual variation term  $\varepsilon_{t,i}$ . In such a case, there is no confounding of passage mortality estimates. However, in the special case where there are systematic differences between the year-effects of up-river stocks and those of mid-to-lower river stocks (i.e. differences significantly greater than those within each stock group), then there is confounding of the parameter estimates for the up-river passage mortality and the region year-effects. At present there is little support for this special case scenario, but it serves to highlight the importance of obtaining information to accurately assess among-group differences in ocean survival (see Discussion).

### Passage Mortality

The emphasis of this study was the estimation of passage mortality experienced by chinook populations of the Snake River subbasins up to the John Day dam. Furthermore, emphasis was given to estimation of that mortality for each of the last twenty brood years (1970–1989). A two-level parameterization scheme was employed in two alternative ways. We separate dam mortality into two categories, as indicated by the 'X's and 'Y's in Table 5-2, and refer below to 'X-type' and 'Y-type' dam mortality. In the first scheme, we assume that  $m_{t,i}$  in Equation [1] is computed by:

$$m_{t,i} = X * n + \mathbf{m} \quad [5]$$

where  $n$  is the number of X-type dams in Table 5-2 (John Day to Bonneville after 1970),  $X$  is the dam passage mortality associated with each X-type dam, and  $\mathbf{m}$  is the net dam passage mortality from the Snake River subbasins to John Day dam ('Y' dams in Table 5-2), expressed as an instantaneous mortality rate for brood years  $\geq 1970$ . The  $\mathbf{m}$  term is a "net" effect mortality estimate because it reflects the overall impacts of dam passage over the complete life cycle, including direct losses due to trauma at the point of dam passage, increased "natural mortality" owing to longer smolt residence time in dam reservoirs, latent mortality due to a weakened condition of smolts, and the benefits or detriments of transportation by barge of some Snake River smolts down-river to below the Bonneville dam.

The first level of parameterization (number of X's in each row of Table 5-2 below) treats mortality as a process proportional to the number of dams passed by a salmon during their transit to the ocean, excluding those dams and/or populations treated in the second level. At the second level of parameterization (Y in Table 5-2) the incremental mortality experienced by upstream stocks is estimated by  $\mathbf{m}$ , which we assume is  $\geq 0$ . Mortality in any given year for any given population is obtained in Table 5-2 by adding the number of Xs ( $n$  in Equation[5]) plus a second level annual term for  $\mu_t$  provided at least one Y is listed. The Warm Springs stock, for example, would currently have  $n=2$  (Bonneville and Dalles dams) and  $\mu_t = 0$ . Johnson Creek, however, would have  $n=3$  (Bonneville, Dalles and John Day) and an estimate of  $\mu_t$  over five dams (McNary, Ice Harbor, Lower Monument, Little Goose and Lower Granite). A symbol is first listed on the diagram for a given dam for the year of initial service (lagged two years to standardize to brood year).

The weakness of the first parameterization scheme is that all the Xs are the same; there is no variation among dams or years. In the second parameterization scheme, we utilize estimates of passage mortality from the CRISP and FLUSH models to create more flexible estimates of mortality for the dams and years with Xs in Table 5-2. This is done in one of two ways. The first approach is to replace  $(X*n)$  in Equation [5] by a passage mortality proportional to that generated by the CRISP/FLUSH models for the X-dams only, that is:

$$m_{t,i} = q * m\text{-from-model}_{t,i} + \mathbf{m} \quad [6]$$

where the "q" is a proportionality constant (estimated in the MLE procedure) to be multiplied by the passage mortality from CRISP/FLUSH for X-type dams. Inclusion of the proportionality constant allows the estimation algorithm to include time and spatial trends in the given passage models (that is, mimic the pattern of mortality changes in CRISP and FLUSH, scaling up or down as required) without constraining the estimated mortality to be the actual value predicted by the passage models. The first approach is used for models 5-8 in the Results (Table 5-4). The second approach is to eliminate  $\mu_{t,i}$  from Equation [6], and estimate  $m_{t,i}$  for all the dams passed by a given stock using the CRISP / FLUSH estimates, either with the proportionately constant  $q$  (models 9-12 in Table 5-8) or using the actual CRISP / FLUSH estimates (i.e.  $q = 1$  in Equation [6]; models 13-16 in Table 5-8).

The CRISP/FLUSH model estimates include both dam-induced effects within the river system plus a background level of "natural" mortality within the river passage. Some natural mortality would be expected even without any dams along the Columbia River, likely in the 0%-20% range for passage between John Day dam and Lower Granite reservoir (see Chapter 6). Thus, some reduction of the CRISP/FLUSH estimated mortalities (in the 0.0 to 0.2 range) would be needed to make them strictly

comparable to the  $m_{t,i}$  estimates, which only consider the *incremental* mortality caused by downstream passage.

**Table 5-2:** Types of passage mortality estimates. (For the simplest models X = fixed estimate of mortality/dam; Y = annually varying estimates of mortality due to passage through 5 dams; BON = Bonneville; TDD = Dalles; JDA = John Day; McN = McNary; IHR = Ice Harbor; LOMO = Lower Monumental; LGS = Little Goose; LGR = Lower Granite). More complex models replace the “X”s with estimates either proportional or equal to CRiSP/FLUSH mortality estimates.

Brood Year	D A M S							
	BON	TDD	JDA	McN	IHR	LOMO	LGS	LGR
1952 - 1954	X			X				
1955 - 1958	X	X		X				
1959 - 1965	X	X		X	X			
1966	X	X	X	X	X			
1967	X	X	X	X	X	X		
1968 -1969	X	X	X	X	X	X	X	
1970 - 1972	X	X	X	Y	Y	Y	Y	
1973 - present	X	X	X	Y	Y	Y	Y	Y

Model (1) together with equation (3) for the  $a$  coefficients contains the following parameters that need to be estimated: these are the fundamental parameters  $d_t$ ,  $a_i$ ,  $b_i$  and  $m_{t,i}$ , and the “true” spawning population abundance’s (in logarithms given by the  $x_{t,i}$ ).

### Likelihood Function and Bias Correction Procedure

The complete likelihood function is made of two parts: the normal errors for the  $e_{t,i}$  and the normal errors for the  $\epsilon'_{t,i}$ . To obtain maximum likelihood estimates (MLE) for all parameters in the complete likelihood problem, one minimizes the negative log-likelihood function, which reduces to minimizing the following sum-of-squares:

$$SSQ = \sum e_{t,i}^2 + I_e \sum e'_{t,i}^2 \quad [7]$$

where the variance ratio  $I_e = \frac{s_e^2}{s_e'^2}$ . Spawner-measurement errors appear to be rather low; a comparison of redd counts to weir counts of females for selected Idaho streams (Lemhi River, primarily) show a coefficient of variation of about 24% in the redd measurements (Appendix 5-1). The likelihood function for the model which includes regional year-effects (i.e. Equation [1] with Equation [4] for  $a_{t,i}$ ) contains an additional sum of squares term:

$$SSQ = \sum e_{t,i}^2 + I_e \sum e'_{t,i}^2 + I_{region} \sum [d_{t,region1} - d_{t,region2}]^2 \quad [7a]$$

where:

$$\begin{aligned}\lambda_{\text{region}} &= \text{variance ratio similar to } \lambda_{\epsilon} \\ \delta_{t,\text{region1}} &= \text{year-effect for mid-lower river stocks} \\ \delta_{t,\text{region2}} &= \text{year-effect for up- river stocks}\end{aligned}$$

The maximum likelihood estimates were obtained by application of a Marquardt nonlinear minimization routine.

A Monte Carlo bootstrap procedure (Efron 1979, 1981) was then applied to the maximum likelihood estimates to produce median unbiased estimates. The bias appears to be rather minor in the results reported later (usually less than 10% of the estimated quantity) and thus a first-order procedure, such as the bootstrap procedure, appears to be a reasonable approach in such cases, although the method may not work well with other data sets which exhibit larger bias (C. Walters, J. Collie; personal communication). The procedure involves construction of a number of simulated data sets with variances obtained by the MLE of original data. In our case we use a parametric bootstrap method in which normally distributed error terms are added to simulated populations in model (1); the variance of the error terms was set equal to the unbiased variance estimate,  $SSQ/(n-p)$ , for the no-spawner-measurement error model ( $n$  is the number of data points and  $p$  is the number of parameters estimated). A similar variance estimate was used in the spawner-measurement error model, except that  $SSQ$  components were partitioned to obtain the recruitment error variance ( $\epsilon_{t,i}$  in Equation [1]). The spawner-measurement error variance ( $\sigma_{\epsilon}^2$  in Equation [1]) was set to  $(.24^2)$  based on actual measurements of spawner measurement error (Appendix 5-1). MLE-estimated variance for spawner measurement error was actually lower than the measured value. Median estimates from the bootstrap trials provide an estimate of bias for the original MLEs. These bias estimates were used as an additive correction to obtain median unbiased estimates; the same additive correction was applied to the bootstrap percentile estimates. In the bootstrap trials, a total of 500 simulated data sets was generated for two of the better models

### Ranking Alternative Models

Two criteria are evaluated for each of the model fits to data, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Akaike 1973, Kass and Raftery 1994). These are both measures of the relative fit of models to data with better model fits corresponding to lower AIC and BIC scores. The AIC, given, by the equation

$$\text{AIC} = -2\ln(\text{Likelihood}) + 2p \quad [8]$$

and the BIC, given by the equation

$$\text{BIC} = -2\ln(\text{Likelihood}) + p*\ln(k) \quad [9]$$

are both based on asymptotic properties of the likelihood function ( $p$  is the number of parameters estimated and  $k$  is the number of “observations,” where we count observation number by the number of  $SSQ$  components. In Equation [7], for example, there are twice as many  $SSQ$  components as data points. Note that both the AIC and BIC use twice the log-likelihood and, hence, both are also related to the asymptotic likelihood ratio tests in which nested models can be tested via classical hypothesis testing (due to the fact that twice the log-likelihood is asymptotically a chi-square statistic (Mood, Graybill, and Boes 1974). Given two models, say M1 and M2, where M2 is the same as M1, except that M2 has  $k$  additional parameters estimated, then we can test the null hypothesis that the simpler M1 is true versus the

alternative that the more complex M2 is true by forming a rejection criterion based on a chi-square statistic with  $k$  degrees of freedom. Both the AIC and to a much greater extent the BIC correspond to a very small rejection criterion (probabilities below .005) for alternative nested models with 21 or more additional parameters in the larger model (as we have in this paper). That is, neither the AIC nor BIC criteria will reject a simple model in favor of a more complex model unless there is a very substantial improvement in SSQ (one significant at  $p < 0.005$ ).

There are differing perspectives on which criterion should be used to select models, or whether specific models should in fact be “selected” in preference to others. The AIC is a conservative method of ranking such nested models. However, there are arguments that AIC results in selecting models with too many parameters, and that the BIC is more appropriate because it forms an approximation to Bayes factors, (the ratio of probabilities of competing models) which is especially important when forecasting is to be done with the selected model. There are counter-arguments to that conclusion, as the BIC has been found to select a simpler incorrect model over the correct more complex model, and that theoretical justification for BIC over AIC depends on the details of changes in the model as  $n$  increases asymptotically. From a Bayesian perspective, Gelman *et al* (1996) argue against model selection altogether, as the alternative models represent alternative hypotheses, and that we would not want to discard the alternative hypotheses through model selection. They argue that what is needed is a general enough model that contains the alternatives as special case solutions so that posterior probabilities calculated from such a general model reflect the relative credibility of alternative hypotheses. A reasonable compromise would be to only reject models which are grossly less supported by AIC and BIC, but reserve judgment on the remaining ones.

#### 5.4 Testing of Models with Simulated Data

The purpose of testing the models is to examine the biases that arise from using a model with the wrong set of assumptions. If the model is consistent in assumption with the data then, as long as the bias is not too large, the bootstrap procedure described above will adjust the estimate to correct for bias. In particular, we wanted to see if the model using Equation [2], where most of the populations shared a common Ricker  $a$  parameter, would perform well even if simulated data contained the assumption that all the  $a$ 's differed. We also wanted to examine the effect of ignoring spawner-measurement error in the 25% coefficient-of-variation range. Therefore, we generated simulated recruitment using the model in Equations [1] and [3], together with an assumed passage mortality  $m_{t,i}$ , and then attempted to estimate this mortality using more simplified versions of the model (e.g. constant  $a$  for all streams, no measurement error, ‘X’ dam mortality set to 0).

Testing of the models involved the construction of a simulated data set and then MLE estimation of parameters for a given model with that data set. The process of simulation/estimation was repeated 100 times in a Monte Carlo procedure. Results are summarized in Table 5-3 below, categorized by model type and simulated data type; the categories do not enumerate all possible combinations but it does show model performance for a wide range of combinations. The range of alternative models was limited to those of the form given by the Ricker model in Equation [1], but with either spawner-measurement error ignored (or not), or with separate Ricker  $a$  values estimated (or not), and with the ‘X’ dam mortality (Table 5-2) estimated (or set to 0).

Simulated data were designed to mimic some variations on the Chinook salmon/ Columbia River system. The simulated data uses the same number of years and populations as given in an earlier version of the spawner and recruit data set (the earlier version covered the same stocks and years as the present one with the exception that there was no 1990 brood year data and the John Day stocks were combined as a single stock). In the simulation, dams operate for the same number of years and areas as in the actual system, but

dam mortality is set to either 0.01, 0.10, or 0.15 per dam. Thus the total mortality for passage through the eight dams present from brood year 1973 is 0.08, 0.80, or 1.20 respectively. Population simulations were initialized with a set of fundamental parameters found from one MLE set of parameter values, including population-specific  $a_i$ ,  $b_i$ , year-effects  $d_t$  (except in limited runs where we set  $d_t = 0$ ), and “true” spawning population abundances set equal to the observed  $S_{t,i}$ . (These parameter values and  $S_{t,i}$  were taken from the earlier data set.) Simulated spawner-measurement error ( $\epsilon'$  in Equation [1a]) was set to a 24% coefficient of variation. Recruitment error  $\epsilon$  in Equation [1]) was set to a 35% coefficient of variation. Both  $\epsilon$  and  $\epsilon'$  were drawn from normal distributions for each simulation. As in the actual MLE runs (Section 5), this test did not incorporate the feedback of recruits becoming spawners; the observed spawning data were considered to be true regardless of the assumed dam mortality.

A useful summary statistic with which to view the results of the simulated trial is the estimated bias in dam mortality — the total passage mortality from the Snake River subbasins through all eight main-stem dams. Table 5-3 refers to this total passage mortality as  $m_{t,10}$ , the value for the tenth stock (Marsh Creek; Table 5-2). The tenth stock was randomly selected as  $m_t$  is the same for stocks 7-13. The estimated bias is assessed using the median of all yearly estimates of  $m_{t,10}$  for the years where all eight dams were operational (1973-1989 brood years). That is, we calculated the average of the 100 MLE estimates of  $m_{t,10}$  for each year and then compared the median of those averages to the true value (i.e. the dam mortality assumed in the simulation) to calculate bias. A positive bias indicates that the estimation procedure overestimates mortality compared to the true value. The standard deviation for  $m_{t,10}$  was also calculated from the 100 simulated/estimation trials; the median of all yearly estimates is included in Table 5-3.

The trade-off between bias and variance is clearly illustrated above in comparisons of scenarios (7,9) or (8,10), where lower variance, but higher bias results are obtained by making the incorrect assumption that the Ricker  $a$  parameters are the same for populations 2-11. Inclusion of spawner measurement error slightly deteriorates performance, as seen by comparison of (7,8) or (9,10). The bias appears relatively constant across a wide range of true dam mortality, as seen by comparison of (7,11,13) provided first-level dam mortality is an estimated parameter. If first-level dam mortality is set to zero then the bias is negative, rather than positive, as seen by comparison of (3,7), and lower variances can be offset by very large biases, as seen by comparison (1,9). With first-level dam mortality set to zero in spite of significant total 8-dam mortality (1, 3, 5), the procedure is unable to generate enough mortality from the 5 “Y” dams in Table 5-2, and therefore shows a significant negative bias (underestimate).

**Table 5-3:** Example results of model testing. Parameter  $m_{t,10}$  is estimated total passage mortality through 8 main-stem dams.

	Model Estimated			Simulated Data Type		Parameter Estimate	
	Different $a$ ?	Spawner-Measurement Error	“X” Dam Mortality Set to 0?	Total 8 Dam Mortality	Year-effect Included?	Bias $m_{t,10}$	Standard Deviation of $m_{t,10}$
1.	Y	N	Y	0.8	N	-0.53	0.22
2.	Y	N	Y	0.08	N	0.02	0.16
3.	N	N	Y	0.8	N	-0.19	0.25
4.	N	N	Y	0.08	N	0.09	0.17
5.	N	N	Y	1.20	N	-0.34	0.24
6.	N	N	Y	0.8	Y	-0.19	0.25
7.	N	N	N	0.8	Y	0.12	0.27
8.	N	Y	N	0.8	Y	0.16	0.28

9.	Y	N	N	0.8	Y	0.08	0.44
10.	Y	Y	N	0.8	Y	0.14	0.51
11.	N	N	N	0.08	Y	0.12	0.19
12.	Y	N	N	0.08	Y	0.10	0.19
13.	N	N	N	1.20	Y	0.12	0.27

A selected number of the scenarios were further examined to investigate bias of the median  $m_{t,10}$  from each simulated year, instead of the average  $m_{t,10}$  of the 100 simulations. The yearly median of those estimates is nearly unbiased for scenarios (9) and (10), as estimated bias is respectively +0.01 and +0.05. No such improvement was found for scenarios (7, 8) in which the Ricker  $a$  parameter is assumed to be the same for populations 2-11; here the medians were just as biased as the means.

Based on the results of the simulation trials, (in particular, the bias associated with estimates of true dam mortality of 0.8 or greater), we conclude that the lowest bias models are the ones which assume no spawner-measurement error, include first-level dam mortality as a parameter to estimate, and can include either of the alternatives that the Ricker  $a$  parameter is different or the same (e.g. rows 7, 9, 13).

## 5.5 Result of Application of the Models to Spring Chinook

A set of 37 spawner-recruitment (S-R) models were applied to the Chinook spawner-recruitment data series covering the thirteen populations described above in Table 5-2. A summary of results from those model runs is given in Table 5-4. The best (that is lowest) AIC scores occurred for models where  $m$  was estimated each of the years (models 1,3,5-8). Among those best AIC scored models, comparable AIC values occur with both of the two alternative parameterizations for first-level dam effects, as described earlier. The best (that is lowest) BIC scores occurred for models with fewer parameters (particularly models 33 and 36), although not those so simple as to omit year-effect parameters. Simple parameterizations for  $m$  were favored by the BIC criterion, either as  $m$  proportional to water transit time (WTT), as in model 36, or with total passage mortality  $m$  proportional to passage model values from CRISP or FLUSH (models 31-34). The WTT variable is the amount of time (in days) that is required for water to pass through the Columbia River system from the head of the Lower Granite reservoir to Bonneville dam, on average during the spring migration period.

**Table 5-4:** Results of applying 37 different models to stock and recruitment data. Shaded models are compared in Table 5-6. "Top models" have AIC and BIC values below the median for all models.

Model Number	Spawner-Measure Error?	Different "a"?	"Top Model"	Distinguishing Features	Total Sums of Squares	Number of Parameters Estimated	Number of Likelihood Components	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Average $\mu$ mortality	Average Ricker "a" Parameter	Production index: $=a-\mu$
1	N	Y	*	Equation (4) passage mortality	111.88	86	406	800.7	1145.2	1.44	2.65	1.22
2	Y	Y		Equation (4) passage mortality	92.14	492	812	1520.9	3833.0	1.26	2.45	1.19
3	N	N	*	Equation (4) passage mortality	120.07	75	406	807.4	1107.8	0.82	1.84	1.02
4	Y	N		Equation (4) passage mortality	96.3	492	823	1553.4	3872.2	0.81	1.84	1.03
5	N	Y	*	X passage mortality factor proportional to CRISP T1	111.49	86	406	799.3	1143.8	1.07	2.12	1.05
6	N	Y	*	X passage mortality factor proportional to CRISP T2	111.47	86	406	799.2	1143.7	1.07	2.13	1.06
7	N	Y		X passage mortality factor proportional to FLUSH T1	113.09	86	406	805.0	1149.6	1.02	1.97	0.95
8	N	Y		X passage mortality factor proportional to FLUSH T2	113.09	86	406	805.0	1149.6	1.02	1.97	0.95
9	N	Y		"m" passage mortality rate proportional to CRISP T1	149.27	65	406	875.7	1136.1	0.59	1.95	1.36
10	N	Y		"m" passage mortality rate proportional to CRISP T2	157.06	65	406	896.4	1156.8	0.02	1.53	1.51
11	N	Y	*	"m" passage mortality rate proportional to FLUSH T1	141.49	65	406	854.0	1114.4	0.99	2.03	1.04
12	N	Y		"m" passage mortality rate proportional to FLUSH T2	152.67	65	406	884.9	1145.3	0.54	1.81	1.27
13	N	Y		"m" passage mortality rates = CRISP T1 values	153.19	64	406	884.3	1140.7	1.00	2.26	1.25
14	N	Y		"m" passage mortality rates = CRISP T2 values	179.72	64	406	949.1	1205.5	0.48	2.04	1.56
15	N	Y		"m" passage mortality rates = FLUSH T1 values	155.02	64	406	889.1	1145.5	1.90	2.51	0.61
16	N	Y		"m" passage mortality rates = FLUSH T2 values	164.07	64	406	912.1	1168.5	1.40	2.29	0.88
17	N	Y		"m" = X*(total number of dams passed)	149.99	65	406	877.7	1138.1	1.44	2.68	1.24
18	N	Y	*	(same as #13 CRISP T1), except exclude 1971 brood year	147.62	63	393	856.3	1106.6	1.00	2.24	1.24
19	N	Y	*	(same as #14 CRISP T2), except exclude 1971 brood year	174.02	63	393	920.9	1171.3	0.48	2.02	1.55
20	N	Y	*	(same as #15 FLUSH T1), except exclude 1971 brood year	141.95	63	393	840.9	1091.2	1.90	2.50	0.60
21	N	Y	*	(same as #16 FLUSH T2), except exclude 1971 brood year	146.86	63	393	854.2	1104.6	1.40	2.27	0.87
22	N	Y		region year effects; "m" as in #1; Lambda(basin)=.01	107.12	118	438	861.9	1343.6	1.43	2.65	1.22
23	N	Y		region year effects; "m" as in #1; Lambda(basin)=1.0	108.64	118	438	868.1	1349.8	1.44	2.65	1.22
24	N	Y		region year effects; "m" as in #13; Lambda(basin)=.01	106.51	96	438	815.4	1207.3	1.00	2.35	1.35
25	N	Y		region year effects; "m" as in #13; Lambda(basin)=1.0	118.62	96	438	862.6	1254.5	1.00	2.32	1.32
26	N	Y		region year effects; "m" as in #14; Lambda(basin)=1.0	126.42	96	438	890.5	1282.4	0.48	2.16	1.68
27	N	Y		region year effects; "m" as in #15; Lambda(basin)=1.0	119.32	96	438	865.2	1257.1	1.90	2.48	0.59
28	N	Y		region year effects; "m" as in #16; Lambda(basin)=1.0	121.62	96	438	873.6	1265.4	1.40	2.33	0.92
29	N	Y		"m" = X*(total number of dams passed); no year-effect	299.48	27	406	1082.4	1190.6	1.88	3.04	1.16
30	N	Y		no year-effect; "m" as in #1, FLUSH T1	289.89	27	406	1069.2	1177.4	1.32	2.02	0.69
31	N	N	*	"m" passage mortality rate proportional to CRISP T1	154.17	54	406	866.8	1083.2	0.55	1.89	1.34
32	N	N	*	"m" passage mortality rate proportional to CRISP T2	167.17	54	406	899.7	1116.1	0.14	1.63	1.49
33	N	N	*	"m" passage mortality rate proportional to FLUSH T1	149.53	54	406	854.4	1070.8	0.71	1.85	1.14
34	N	N	*	"m" passage mortality rate proportional to FLUSH T2	158.23	54	406	877.4	1093.7	0.53	1.77	1.24
35	N	N	*	"m" = X*(total number of dams passed)	158.56	54	406	878.2	1094.6	0.57	1.94	1.37
36	N	N	*	$\mu$ passage mortality rate proportional to WTT	145.82	55	406	846.2	1066.6	0.79	1.78	0.99
37	N	Y		$\mu$ passage mortality rate proportional to WTT	136.28	66	406	840.8	1105.2	1.34	2.52	1.19
				Average across all models	144.59	94	433	909.2	1306.2	1.03	2.17	1.14
				Median value	145.82	65	406	868.1	1145.3	1.00	2.13	1.19

A subset of eight of the models are shown in Figure 5-1, typical of ones for which both a year-effect and a yearly  $m$  parameter were estimated. Four of the models use the first parameterization of passage mortality given above (i.e. equation [5]; constant mortality per project) and four of the models use the second parameterization (i.e. equation [6]; per project mortality proportional to CRISP/FLUSH estimates). The first four models (models 1-4 in Table 5-4) are given by Equation [1], with all possible binary choices to the two model attributes (include/not include spawner-measurement error) and (Ricker  $a$  parameter different for each stock / same for 12 stocks). A value of  $I_e = 2.0$  was used in the spawner-measurement error models because that corresponds approximately to an assumed 24% coefficient of variation in spawner-measurement error (see Appendix 5-1). Varying measurement errors by stream based on PATH participants' judgments of the relative accuracy of spawning enumeration methods (Marmorek 1995) changed the average  $m$  estimate by only 0.006 in earlier tests of one of the better models with preliminary data.

The second group of four models in Figures 5-1 (models 5-8 in Table 5-4) use the second (more variable) passage mortality parameterization, one for each of two versions of the CRISP passage model and two versions of the FLUSH passage model. The passage models are embedded in Equation [1], with the assumption of no spawner-measurement error and that the Ricker  $a$  parameter is different for each of the stocks. The final model, number 37 of Table 5-4, is included in Figure 5-1 because of its low BIC score and because it contains a simplified passage model:  $m$  proportional to WTT. This model has a slightly higher BIC score than model 36, but was chosen because it includes different Ricker  $a$  parameter values (shown in the simulation tests to perform more accurately).

All eight S-R models gave similar estimates of time trends in both  $m$  (passage mortality from Snake River sub-basins to John Day Dam) and in  $d_t$  (year-effect parameter) (Figure 5-1). The time trend in  $m$  is one of a cyclical high to moderate mortality throughout the 1970s followed by a four-year period of low mortality (1980-1983) then a return to a stable but slightly cyclical mortality through the latter half of the 1980's and an upward spike in 1990 (Figure 5-1a). Results are also given in Figure 5-1a on the average of the CRISP and FLUSH mortality estimates for their four configurations (discussed in detail below). As seen in Figure 5-1, the average passage model estimates of  $m$  generally pass through the  $m$  estimates of the eight S-R models with three notable exceptions: the spiked mortality indicated by the passage models for brood year 1971 is not present in any of the S-R models; secondly, the mortality indicated by the passage models for brood years 1980-1983 does not show the precipitous reduction in mortality (increase in survival) indicated in all the S-R models; and thirdly, the increased mortality in all eight models in 1990 is above the average passage model. (A more detailed examination of the four-year high survival result in 1980-83 was made in of our previous report. We had wanted to see if the reduction in mortality for the upriver populations was due to real increases in recruitment or just due to decreases in spawning population; both occur. The decrease in spawning populations was also observed on the lower river populations, but they did not experience the increase in recruitment that was seen in upriver populations.) A moderate long-term cycle is shown in Figure 5-1b for estimated "year-effects" indicating a shift in climate mortality regimes from a generally high positive anomaly sequence for the 1952-1968 brood years followed by generally negative anomalies for the 1970-1989 brood years.

### Confidence Intervals for $m$ and $d$

Results of the bootstrap procedure are shown in Figure 5-2 for two of the better models (model numbers 1 and 3), as judged from both the simulation trials and the two information criteria in Table 5-4. Emphasis is given to the top panel of that Figure because of its lower bias in the simulation trials, but the other panel shows consistent results. Bias of the MLE was generally small as an estimator of median  $m$  (range from +.008 to +.06 for the two models' time-averaged  $m$ ). The MLE model's estimates of " $a$ " and " $X$ " are correlated; runs with higher productivities (e.g. " $a$ "s estimated independently) require higher estimated

dam mortalities to match the spawning estimates. Bias-corrected estimators are used for all model results given in this Section. Other results are listed in Table 5-5 below. (Appendix 5-I of our previous draft included several graphs from the second model in Table 5-5, including graphs by population of predicted and observed recruitment and spawners displayed in several alternate formats. We didn't include an updated version because results are quite similar to those given earlier for common years and stocks.)

**Table 5-5:** Estimated parameters for Ricker model for four model types. Also listed are 90% confidence intervals for the first model type and first-level per dam mortality rates of "X" type as described in the text.

Model Type						
Measurement error	N	Y	N	Y	N	
"a" differ, some	Y	Y	N	N	Y	
Subbasin	Ricker "a" values				90% Confidence Interval	
Wind R	1.40	1.36	1.36	1.34	1.05	1.74
Klickitat	1.72	1.71	1.88	1.88	1.41	2.00
Warm Sp	2.88	2.73	1.88	1.88	2.39	3.31
John Day Mainstem	2.49	2.39	1.88	1.88	2.01	2.96
John Day Mid-fork	2.19	2.07	1.88	1.88	1.74	2.65
John Day North-fork	2.23	2.06	1.88	1.88	1.64	2.78
Imnaha	3.08	2.74	1.88	1.88	2.47	4.07
Minam	3.15	2.90	1.88	1.88	2.61	4.12
Bear Val	3.10	2.81	1.88	1.88	2.50	4.05
March Cr	3.01	2.69	1.88	1.88	2.42	3.95
Sulphur C	3.29	2.96	1.88	1.88	2.72	4.22
Poverty Fl	2.92	2.70	1.88	1.88	2.34	3.92
Johnson F	3.00	2.75	1.88	1.88	2.43	3.98
Subbasin	Ricker "b" values (x1000)					
Wind R	3.52	3.34	3.50	3.46	2.16	4.90
Klickitat	2.65	2.78	3.12	3.60	1.95	3.56
Warm Sp	1.15	1.01	0.39	0.30	0.66	1.66
John Day Mainstem	2.96	2.82	2.71	2.62	1.95	3.78
John Day Mid-fork	1.41	1.21	1.70	1.59	0.93	1.86
John Day North-fork	0.46	0.40	0.55	0.52	0.22	0.65
Imnaha	0.69	0.53	0.65	0.57	0.46	0.90
Minam	1.22	1.07	1.08	0.95	0.87	1.54
Bear Val	0.60	0.49	0.53	0.47	0.34	0.81
March Cr	1.01	0.70	1.00	0.85	0.52	1.40
Sulphur C	1.94	1.51	1.46	1.23	1.36	2.63
Poverty Fl	0.66	0.65	0.70	0.71	0.48	0.86
Johnson F	1.95	1.80	1.98	1.89	1.33	2.53
Dam mortality rate						
"X" per dam	0.264	0.244	0.033	0.050	0.138	0.442

### Comparison of $m$ with Upriver-Downriver Differences in $\ln(r/s)$

Empirical examination of some of the features of the spawner-recruitment data gives a better understanding of the time trends in mortality estimates. In Figure 5-3 we show ad-hoc estimates of average standardized mortality of recruits for upriver and down-mid river populations. The ad-hoc estimates were obtained by subtracting the time series average  $\ln(R/S)$  from each year's  $\ln(R/S)$  for each population and then averaging those across populations in the upriver (or alternatively down-mid river) sub-basins. This kind of ad-hoc technique is imperfect, but it does show some interesting features. It also permits a rough comparison between our analyses and those of Chapter 3. The  $m$  passage mortality from our "best" S-R model nearly mirrors the difference between average  $\ln(R/S)$  for upriver versus down-mid river populations in most years. For example, the spikes upward in  $m$  for recruits born in brood years 1972 and 1975 mirror the spikes downward in upriver survival relative to the average survival experienced during those years by the down-mid river populations. The low estimates of  $m$  for 1980-1983 brood years reflect the better than average survival of upriver stocks compared to the survival of low-mid river populations (recall that  $m$  is kept  $\geq 0$ , otherwise it would be negative during this period). The very large upward spike in passage mortality indicated for brood year 1971 recruits by the passage models (Figure 5-1) is not supported by the changes in survival of the upriver populations relative to the down-mid river populations. The upriver recruits of brood year 1971 had a lower survival, but not as low as experienced by the 1972 or 1975 brood years. Details on survival trends in each of the populations show that brood year 1972 was the poor year, not 1971, especially for the Bear Valley, Sulphur Creek, Poverty Flats, and Johnson Flats populations (Appendix 5-I, previous report). The ad-hoc estimates fail to account for the  $m$  estimates obtained in the latter 1980s, but they do reflect the drop in survival of up-river stocks in 1990.

### Comparison of $m$ with CRISP and FLUSH Based Passage Mortalities

Two transport assumptions were applied to CRISP and FLUSH which led to four estimates of passage mortality (Figure 5-4). The transport survival assumptions in Figure 5-4 are based on the ratio of survival of transported fish to control fish left in river (transport:control ratio or TCR) in selected years of experiments, multiplied by the in-river survival estimated by each passage model for those years. Transport model 1 (T1) assumes that survival varies with water transit time; the slope is anchored by TCR values of 1.0:1 in 1986 (a high flow, low WTT year) and 3.0:1 in 1977 (a low flow year). Transport model 2 (T2) assumes fixed survival, calculated based on a TCR of 1.6:1 from 1986. FLUSH clearly estimates much lower transport survival than CRISP (Figure 5-4).

The passage models provide estimates of passage mortality for Spring Chinook for their entire seaward migration. We used those passage mortalities to estimate  $m$ , as described in Section 5.3. We completed several sets of four model runs (corresponding to the two transport models for each of the two models CRISP and FLUSH). The three passage mortality scenarios (CRISP T1, FLUSH T2, FLUSH T1) are more similar to each other than the CRISP T2 scenario, as seen in Figure 5-5 and by results listed in Table 5-4. In each group of four scenarios listed in Table 5-4 where  $m$  is based on one of those passage models, (that is models 9-12, models 13-16, models 18-21, models 25-28, models 31-34), the CRISP T2 scenario had the worst (largest) AIC and BIC score. Closer inspection, however, shows that the year of comparison affects which of the passage model scenarios most closely match the estimates of  $m$  from one of the better empirical models (such as model number 1 where annual  $m$  values are estimated). The CRISP T1 and T2 models (numbers 5 and 6 of Table 5-4) produced the best AIC score of all models when CRISP estimates are used as scalars to estimate 'X-type' dam mortality (i.e. John Day, the Dalles, Bonneville), together with the MLE estimate for other dams (Equation [6]). Note that transportation is not part of the passage model mortality estimates for the 'X-type' dams.

Table 5-6 provides some quantitative measures of comparison with the S-R model number 1 of Table 5-4 (shown in Figure 5-5). The CRISP T1 estimates produce the best overall fit in terms of lowest  $SSQ$  in criterion (5), but that result can be altered (criterion 7) to favor the FLUSH T1 estimates if brood year 1971 is not fitted — a year in which estimates are strikingly different from those of the S-R model (Figure 5-5). The CRISP T2 estimates of  $m$  are not within the 90% CI of the  $m$  estimates from the S-R model as often as the other passage model estimates and it has considerably larger absolute deviations (criteria 1,2). However, that result has to be tempered by the fact that CRISP T2 produces closer  $m$  estimates during the lower mortality brood years of 1980-1983. Average passage mortality obtained from the FLUSH passage models are larger than the average  $m$  from the S-R model while the CRISP passage models are smaller than the average  $m$  as seen in criterion (4). The average deviation is smallest for the CRISP T1 and FLUSH T2 models. We have not adjusted the CRISP or FLUSH estimates by subtracting pre-dam natural mortality (in the 0.0-0.2 range), which could alter comparison results slightly (likely improving the fit of FLUSH to the S-R data, and reducing the fit of CRISP).

**Table 5.6:** Various indicators of passage model performance. (Models 13-16 vs. Model 1 of Table 5-4 for rows 1-7; models 18-21 for row 8.)

Criterion	CRISP T1	CRISP T2	FLUSH T1	FLUSH T2
1. % of $m$ within 90% CI	52%	33%	57%	57%
2. Average absolute deviation from median estimate	0.75	1.12	0.73	0.72
3. Median absolute deviation from median estimate	0.61	1.07	0.58	0.63
4. Average deviation from median estimate	-0.34	-0.87	0.55	0.06
5. $SSQ$ given passage model $m$ but other parameters estimated	153.19	179.72	155.02	164.07
6. AIC	884.3	949.1	899.1	912.1
7. BIC	1140.7	1205.5	1145.5	1168.5
8. $SSQ$ as in (5.), omitting 1971 brood	147.62	174.02	141.95	146.86

Regional year-effects were investigated in model numbers 22 to 28. All of the BIC scores are substantially poorer for those regional models as compared to ones without region -specific year-effects (Table 5-4). The AIC scores are also poorer for models 22 and 23 than for model 1, one of the better empirical models. However, the CRISP and FLUSH based models do show lower AIC values when region-specific year effects are included, perhaps because this gives more flexibility to accommodate upstream-downstream differences in survival (models 24-28 vs. 13-16). Passage mortality rates are very similar between those two types of models, whereas the regional models estimate a slightly larger average Ricker  $a$  parameter. However, year-effects are strongly supported for inclusion in the models, as seen by the high AIC and BIC scores for model numbers 29 and 30 in which no year-effect was included.

Using model 1, we examined estimates of  $m$  (up-river net passage mortality), and  $d$ , (year-effects) for correlations with other variables. In Figure 5-6, we show a graph of  $m$  estimates versus WTT; there is a significant positive correlation ( $r = .58$ ,  $p < 0.01$ ) between those variables. The correlation is still significant without the influential 1975 brood-year ( $r = 0.45$ ,  $p < 0.05$ ). We separated the 1970-79 and 1980-90 periods to assess if differences in power system management or other factors may have affected the correlation of WTT and  $m$ . The 1970s showed a weaker correlation ( $r = 0.58$ ,  $p = 0.08$ ; without the 1975 datum,  $r = 0.06$ ,  $p = 0.87$ ) than the later period ( $r = 0.68$ ,  $p = 0.02$ ). However, an Analysis of Covariance with WTT as the covariate showed no significant difference between the 1970s and 1980s ( $p = 0.623$ ). Some pairs of years with very similar WTT (e.g. 72-70; 74-80; 89-83; 90-86) had very different values for  $m$ . This indicates that other factors besides WTT influenced the value of  $m$  (Figure 5-6a) in these years.

The  $d$  year -effect reflects common factors affecting survival of all stocks. We therefore selected two general climate indicators applied by Charlie Paulsen in Chapter 4, the NPI and the Upwelling Index, to see if they were correlated with  $d$ . The NPI showed low correlations in all cases when matched to the first, second, and third ocean winter (i.e.  $r = 0.08, 0.26, 0.33$  respectively). The Upwelling Index was also not correlated with  $d$  ( $r = -0.09$ ). Interestingly, Dr. Randall Peterman (unpub. data) found that the standardized residuals of Bristol Bay sockeye survivals were negatively correlated with  $d$  ( $r = -0.51$ ). This is consistent with the hypothesis that ocean conditions in the Northeastern Pacific vary inversely with those off the coast of Oregon and Washington.

Correlations are not significant in either the mid-panel of Figure 5-6 ( $d$  versus WTT;  $r = -0.15$ ) or Figure 5-7 ( $m$  versus  $d$ ;  $r = -0.24$ ). The absence of any correlation between  $d$  and WTT suggests that WTT is not a good indicator of whatever common factors are driving the year effect. The lack of correlation between  $\mu$  and  $d$  (Figure 5-7) counteracts the potential concern that the estimates for these two parameters might be unstable and “trading off” against each other. Rather, the passage mortality and year effects appear to be largely independent. The only obvious pattern in Figure 5-7 is that the brood year with the best year effect (1983) also had a low passage mortality  $m$  and the brood year with the worst year effect (1990) also experienced high passage mortality.

The wide range of alternative hypotheses regarding the magnitude of up-river net passage mortality  $m$  and Ricker “a” parameter are supported by the results in Table 5-4. A range in average  $m$  from 0.55 to 1.90 (mean of 1.09) occurs for the “top models” in Table 5-4, where we define the “top models” to be those models for which both the AIC and BIC are below the median for all 37 models. This corresponds to a range of 42-85% [ $100 * (1 - e^{-m})$ ] mortality due to downstream passage from Lower Granite to John Day dams, over the entire life cycle (mean of 66%). There were 12 models which fit in the top model category: models 1,3,5,6,11,18,20,21,31,33,36, and 37 of Table 5-4. These models’ average Ricker “a” values range from 1.78 to 2.65 (mean of 1.92). They have a smaller range in average production index  $a - m$  from 0.60 to 1.34 (mean of 1.06). The smaller range in average production reflects the correlation between estimates of  $m$  and  $a$ , which is present not only in the results of Table 5-4, but also was seen in the bootstrap simulations done for models 1 and 3. Higher passage mortality estimates require higher production values to match observed patterns in S and R (Figure 5-8).

## 5.6 Discussion

### Implications of Retrospective Analysis

We were encouraged by how closely the most empirical of the MLE passage models (model 1 of Table 5-4) gave passage mortality estimates (based on S-R data) that generally matched those produced by the more mechanistic passage models (particularly CRISP T1, FLUSH T1, and FLUSH T2). In general, those MLE estimates of mortality are intermediate between the CRISP and FLUSH estimates. Better agreement between the mechanistic and MLE passage models was achieved by building a model in which one scales the mortality estimates from CRISP and FLUSH. On the other hand, even the simple explanatory variable, water transit time, could be used to model passage mortality in a high performance model (as in models 36 and 37 of Table 5-4).

The influence of in-river studies on the 1971 brood year (conducted on migrants in 1973, a low-flow year) has been profound in the debate over the effects of dams on mortality of smolts. Yet our results show that the 1973 study results are not consistent with observed returning recruits from the 1971 brood year. If we

assumed for a moment that the high mortality rate of FLUSH for the 1971 brood year was correct (i.e. a passage mortality  $m$  which is 3.4 above the  $m$  given by the S-R model) then we must have had an exceptionally strong year-class for the brood year 1971 in terms of their natural survival. In particular, we would require that observed recruitment survival rate,  $\ln(R/S)$ , to be roughly 3.4 above the S-R model's average survival rate (i.e. the rate consistent with the S-R model's average passage mortality). Figure 5-3 gives a rough idea as to how strong that year-class would need to be: adding 3.4 to the standardized  $\ln(R/S)$  for brood year 1971 would place it about 1.0 above the maximum standardized survival for any other brood year for the upriver stocks. On the other hand, the high passage mortality for the 1975 brood year (conducted on migrants in 1977, another low-flow year) is generally consistent between the passage models and the S-R model and also consistent with in-river studies conducted that year.

An alternative explanation for our results is that the decline in survival of up-river vs. down-river natural stocks was due to a marine mortality factor that differentially affected up/down river stocks in a large enough way to coincide with and account for the estimated dam mortality. The models which include regional year-effects (models 22-28) generally performed worse than the empirical models which exclude this alternative hypothesis (e.g. model 1). We have not been able to find any analyses clearly indicating *systematic, regional differences* between up-river and mid-to-low river stocks in ocean distribution or mortality. Coded wire tag data from five spring chinook hatcheries, discussed in Section 4.5 of Chapter 4 (Paulsen 1996), do show significant differences among *all* stocks in ocean distributions, but provide no indication that the differences between up-river and mid-to-low river stocks are larger than the differences within each stock group. PATH participants pointed out a number of other problems with the CWT data analyzed by Paulsen (1996); these were summarized by Marmorek (1996). The multivariate analyses in Chapter 4 (Table 4.6) show some differences between stock groups in the variables which correlate with  $\ln(R/S)$ , but there are also many differences within each stock group (Table 4.1.1). Though it would be valuable to obtain better estimates of among-region differences in ocean distribution and survival of wild spring chinook, there are clearly not enough wild fish remaining to generate sufficient tag recoveries.

The models developed in this paper omit statistical covariance in the measurement errors. We did not attempt to model directly the process by which recruitment estimates were obtained by Petrosky et al (1995), although such modeling is theoretically possible. The advantage of a more complete model for the measurements of recruitment is that we would then have the appropriate error structure to include the statistical covariance structure between recruitment measurement error and the several years of spawner-measurement error to which they contribute. Our opinion at present is that the covariance is likely low for two reasons. First, recruitment measurement errors originate from many important sources other than those induced by errors in escapement counts, including conversion rates that are applied at each dam to convert dam counts to total passage estimates, partition rates to partition passage estimates among the measured and non-measured streams, and harvest removal corrections. Second, Chinook salmon return to the Columbia River to spawn (and hence become part of recruitment) over four or more ages and thus the spawner-measurement error calculated in any given year would be applicable only to a portion of the recruitment from a given brood year.

Other concerns are the covariance in recruitment measurement errors among populations that share nearby sub-basins (such as the Snake River populations) and auto-correlation between recruitments in adjacent years. We did examine the latter and found it had a negligible effect.

One suggestion (J. Anderson, pers. comm.) would be to consider an alternative parameterization of  $m$  that takes the additional explanatory variable, percent of Snake River stocks transported, into account. At present, the approach does not partition  $\mu$  into its component sources (i.e. transportation, dams, reservoirs, other migratory habitat impacts). This is consistent with the PATH Level 2 approach (see Chapter 1). Incorporating a number of additional explanatory variables is an interesting avenue for future research, one which would move the models closer to a Level 3 approach.

## Prospective Analyses

There is clearly uncertainty about the magnitude of passage mortality and of the Ricker “a” parameter, although less uncertainty about net production of the stocks. One way to address that uncertainty is to construct prospective analyses for a single model which encompasses a range of alternative hypotheses about those parameter values. Such a model is Model 1 of Table 5-4 which contains year-specific  $m$  passage mortalities and individual stock Ricker “a” parameters, and is in a virtual tie for best model according to the AIC criterion. Model 1 showed an inherent correlation between  $m$  estimates and “a” parameter estimates in the bootstrap results ( $R=0.90$ ) and thus builds a feature into the analysis that was also present in results in Table 5-4. In addition, that model will need to be generalized (as shown below) to have a stock recruitment curve which encompasses a wider range of hypotheses about density dependence.

The models and results of this paper can be used in the next phase of research on Prospective Analyses. Some changes will need to be made. Bayesian analysis was not necessary for the Retrospective analysis, but such work is critical for decision making. It is also important that the stock recruitment curve used in prospective analyses consider both a dropping right-hand limb (i.e. a Ricker curve) and also depensatory effects. Equation [6] is a method of fitting stock recruitment data which allows for both depensatory and Ricker-like effects. In this equation the  $b$  parameter affects the amount of depensation, “c” a carrying capacity term, and  $g$  affects the right-hand limb of the curve. We also will examine a three-parameter curve recommended by Dr. J. Collie in an earlier review. By obtaining probability distributions for these parameters, the prospective analysis would project future states of populations based on the overall uncertainty in stock recruitment curves. Thus, for many stocks where there are insufficient data to clearly demonstrate either depensatory effects or a dropping right-hand limb, there would be a relatively low probability of those types of stock recruitment curves being selected in the prospective analysis. It is important, however, that the curves used for prospective analyses receive close scrutiny for their biological reasonableness, since the future trajectories of populations may carry them beyond available data.

$$R = \frac{aS^b}{(1 + cS)^g} \quad [10]$$

A first step would be to assess the overall improvement in survival required by different populations to reach certain target levels with a specified level of confidence (see Chapter 6). This analysis will help to set targets for the aggregate effect of survival improvements in different life stages.

The mechanics of generating prospective analyses are not fully developed. However, some steps are clear. Bayesian analysis will produce posterior probability distributions for a number of parameters ( $m$  for each year between 1970 and 1990,  $d$  for each year, and the stock recruitment parameters in Equation [6] (some of which (e.g.  $c$ ) will be indexed by stock)). The Bayesian analysis will generate many thousands of runs, each consisting of a different vector of parameter values. This vector would include not only the four parameters for the stock recruitment curve but also the 20 estimates of “ $m$ ,” and  $d$ , for each of the years included in the analysis. For the prospective analysis, one could draw a vector of stock recruitment parameters that is used throughout a 100-year future simulation. If it is assumed that future passage mortality mimics the 1981-89 period, for example, then one could randomly choose one of the nine “ $m$ ” values associated with the chosen stock recruitment parameter values. In addition, one could randomly draw a process error based on a normal distribution for each year. Different assumptions about future passage survival could be explored using such a framework. Formal decision analysis techniques could be applied, as discussed in the report from the PATH Workshop held in April, 1996 (Marmorek et al. 1996).

## 5.7 References

- Akaike, H.** 1973. Information Theory and an extension of the maximum likelihood principle. *in* Second International Symposium on Information Theory, eds. B.N. Petrox and F. Caski, Budapest: Akademiai Kiado. 267pp.
- Anderson, J., J. Hayes, and R. Zabel.** 1996. Columbia River Salmon Passage Model: Theory, Calibration and Validation. University of Washington, Center for Quantitative Science.
- Beamesderfer, R.C.P., C.E. Petrosky, C.E., H.A. Schaller, and O. Langness.** 1996. Spawner-recruit relationships for spring and summer Chinook salmon populations in several Columbia and Snake River subbasins. PATH report (in preparation).
- Bradford, M.J.** 1995. Comparative review of Pacific salmon survival rates. *Can. J. Fish. Aquat. Sci.* 52: 1327-1338.
- Efron, B.** 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7: 1-26.
- Efron, B.** 1981. Non parametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* 68: 589-599.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin.** 1996. Bayesian Data Analysis. Charpman & Hall, London. 526p.
- Kass, R. E. and A. E. Raftery.** 1994. Bayes Factors. Tech Rep. 254. Dept Statistics, Univ Wash. 56p.
- Marmorek, D.R., D.R. Bouillon, and I. Parnell.** 1996. Plan for Analyzing and Testing Hypotheses (PATH): Results of the Kah-Nee-Ta Workshop on Retrospective and Prospective Analyses, (April 17-19). Prepared by ESSA Technologies Ltd., Vancouver, BC for Bonneville Power Administration, Portland, OR, 52 pp. and appendices.
- Marmorek, D.R.** 1995. Summary of meeting held November 16-17 in Portland, Oregon. Technical memo. Dec. 1, 1995.
- Marmorek, D.R.** 1996. Summary of comments on Paulsen May 23 memo discussing ocean harvest distribution of hatchery spring chinook. Memo. June 19, 1996.
- Mood, A. M., F. A. Graybill, D. C. Boes.** 1974. Introduction to the theory of statistics. McGraw-Hill, New York. 564 p.
- Paulsen, C.M.** 1996. Ocean harvest distribution of hatchery spring chinook. Memo. May 23, 1996.
- Ricker, W.E.** 1976. Review of the rate of growth and mortality of Pacific salmon in salt water, and noncatch mortality caused by fishing. *J. Fish. Res. Board Can.* 33: 1483-1524.
- Walters, C.J, J.S. Collie, and T. Webb.** 1989. Experimental designs for estimating transient responses to habitat alteration: Is it practical to control for environmental interactions? p.13-20, in C.D. Levings, L.B. Holtby, and M.A. Henderson [eds.]. Proceedings of the National Workshop on Effects of Habitat Alteration on Salmonid Stocks. *Can. Spec. Publ. Fish. Aquat. Sci.* 105.
- Wilson, P.** 1994. Spring FLUSH (Fish Leaving Under Several Hypotheses), version 4.5. Draft documentation. July 15, 1994. Columbia Basin Fish and Wildlife Authority. 27 pp.

**Figure 5-1** a. Estimates of system mortality (*m*) above John Day dam by brood year for eight stock-recruitment models (numbers 1-8 in Table 5-4) and a simplified passage model in which *m* is proportional to WTT. b. Year-effect (*d*) estimates for the same nine models.

**Figure 5-2** Bootstrap estimates of confidence intervals for  $m$  and  $d$  for models 1 and 3 in Table 5-4 (Model 1:  $a$  different for all stocks; Model 3:  $a$  same for stocks 2-13).

**Figure 5-3** Comparison of normalized  $\ln(R/S)$  for upstream and downstream populations, and MLE estimates of  $m$ . The value of  $\mu$  is generally proportional to the down-mid river line minus the upriver line.

**Figure 5-4** Comparison of CRiSP and FLUSH transport model survivals. Model T1 assumes that transport survival decreases with water travel time, whereas model T2 assumes constant survival.

**Figure 5-5** Comparison of the MLE median estimates of  $m$  using model 1 (see Table 5-4), with those from four mechanistic passage models. Confidence intervals are also shown, based on the bootstrapping approach described in the text.

**Figure 5-6** a. Estimated passage mortality ( $m$ ) vs. Water Transit Time (WTT). The regression line was fit through all of the data. The second y axis converts  $m$  to an estimated mortality over the complete life cycle [ $100*(1-e^{-\lambda})$ ]. b) Year effect ( $d$ ) vs WTT.

**Figure 5-7** Year effect (*d*) vs. estimated passage mortality (*m*).

**Figure 5-8** Correlation between estimates of Ricker  $a$  parameter (average over all stocks) and 21-year average of passage mortalities ( $m$ ) for all 37 models in Table 5-4.

---

## Chapter 5 Appendix 1

### Measurement Error in Escapement Estimates

Charlie Petrosky

Measurement error in escapement indices contributes variability to both spawner and recruit estimates. In the MLE framework for Level 2 analyses, a specific measurement error term is defined for estimates of spawner escapement. Consequently, an estimate of the magnitude of measurement error is desirable for each of the populations in the analysis. The objective of this document is to summarize information on measurement error in ground redd counts from selected Idaho streams.

If weir counts of females are taken to be accurate measurements of actual spawners, variability in the relationship between redd counts and weir counts can be considered to represent redd count measurement error. Other potential sources of variation (e.g. prespawning loss, misclassification of sex at the weir, number of redds constructed by females) would be subsumed within this error term. Because these data are from populations are not represented as index stocks in the run reconstructions (Petrosky et al. 1995), they constitute an independent estimate of measurement error for stocks to be included in the MLE framework.

#### *Methods*

Data sources include numbers of females counted past weirs and subsequent ground counts of redds for the Lemhi River, 1965-1974 (Bjornn 1978), the upper Salmon River at Sawtooth Hatchery, 1988-1992, and Crooked River, 1991-1992 (Kiefer and Lockhart 1994). These data represent published, readily available observations, and not an exhaustive data search. Cases where weir counts were incomplete, or did not represent spawning escapement, were excluded from the data set (Lemhi River--1964; upper Salmon River--1989-1990).

Ground counts of redds were regressed against the female escapement at the weir. In addition, a ratio was calculated for each observation of redd count to female escapement at the weir; the mean and coefficient of variation of ratios was reported.

#### *Results and Discussion*

Lemhi River weir counts of female chinook, 1965-1974, ranged from 206 to 969, while redd counts ranged from 237 to 786 (Table 1). Recent observations from the upper Salmon River and Crooked River ranged from 5 to 275 females past weirs, and from 4 to 261 redds counted.

Numbers of redds counted corresponded well ( $r^2 = 0.91$ ,  $p < 0.01$ ) with females counted past the weirs (Table 1, Fig. 1). On average, the ratio of redds counted to female escapement was 0.90, with CV = 0.24 (Table 1)

#### *Literature Cited*

**Bjornn, T.C.** 1978. Survival, production, and yield of trout and chinook salmon in the Lemhi River, Idaho. Bulletin Number 27. Forest, Range and Wildlife Experiment Station, University of Idaho. 57 p.

**Kiefer, R.B. and J.N. Lockhart.** 1994. Intensive evaluation and monitoring of chinook salmon and steelhead trout production: Crooked River and upper Salmon River sites. Annual Report 1992. Idaho Department of Fish and Game. Prepared for U.S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife. Project 93-7. Contract No. DE-BI79-91BP21182.

**Petrosky, C.E., H.A. Schaller and R.C.P. Beamesderfer.** 1995. Spawner-recruit relationships for spring and summer chinook populations in several Columbia and Snake River subbasins. Incomplete draft, October 11, 1995. Prepared for PATH Workshop 1.